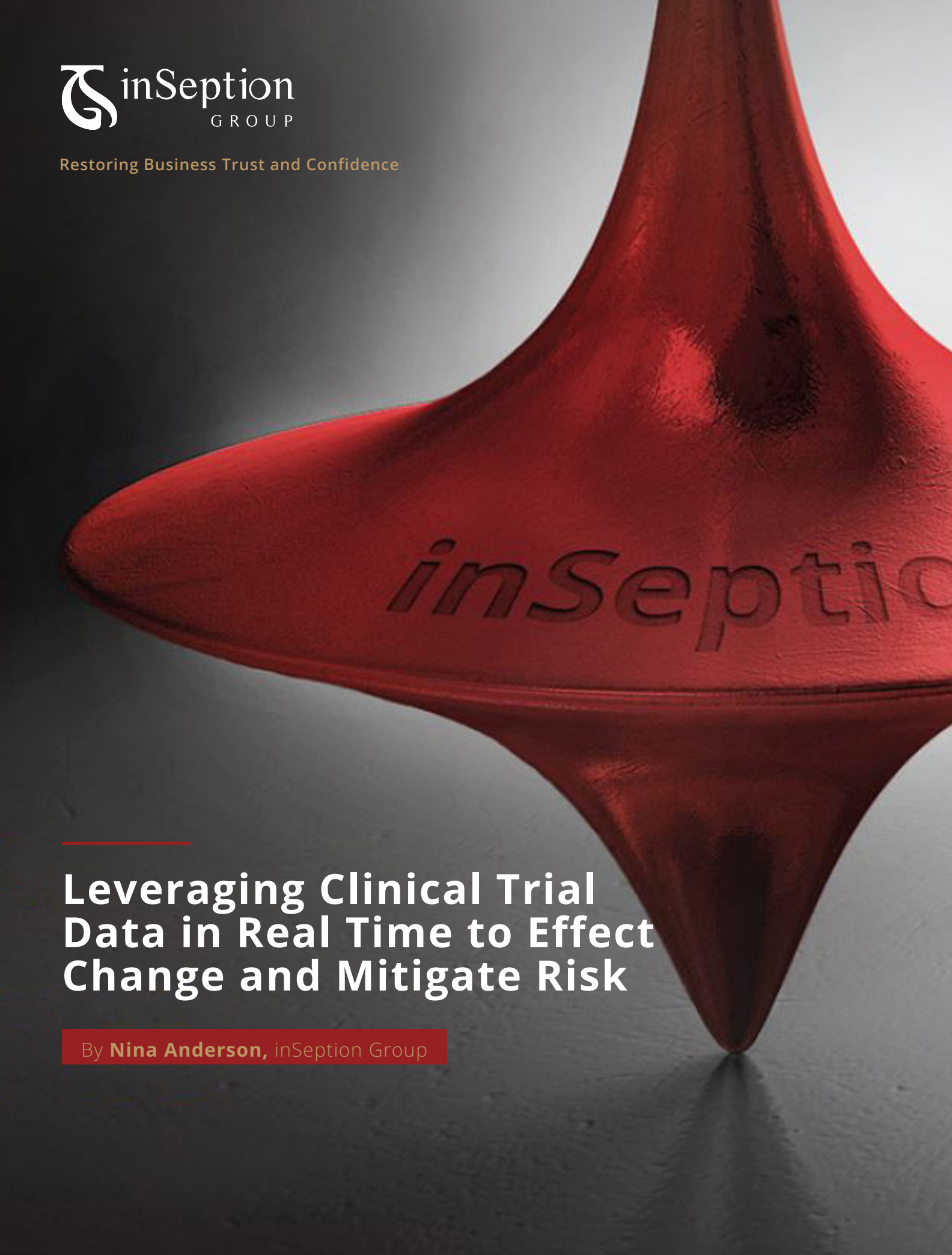
inSeption GROUP

Restoring Business Trust and Confidence



Leveraging Clinical Trial Data in Real Time to Effect Change and Mitigate Risk

By **Nina Anderson,** inSeption Group

Clinical trial data gathering, contextualization, and analysis all can benefit from existing — but underutilized — technologies and resources.

Leveraging Clinical Trial Data in Real Time to Effect Change and Mitigate Risk

By **Nina Anderson,** inSeption Group

Properly applied, data analytics can improve clinical trial data quality, as well as potentially reduce the costs and time investment associated with traditional trial monitoring. But advanced analytics are critically underutilized in the pharmaceutical industry, particularly in early-stage clinical trials, due to concerns over the investment, in-house expertise, or regulatory scrutiny, among other worries.[1]

However, pharmaceutical sponsors face a multitude of challenges that could be addressed through better understanding and application of analytics and data science. These obstacles include collecting high-quality data, managing the volume of data produced, and analyzing that data. Sponsors also are inhibited by the length of time needed to clean and monitor data, plus inconsistent access to meaningful, accurate data visualizations. Through better understanding of these challenges and the solutions available to overcome them, the industry can progress to more efficient and well-managed clinical trials whose results are trusted by regulatory authorities.

SOURCE DOCUMENT VERIFICATION DOES NOT GUARANTEE DATA QUALITY OR ACCURACY

Ensuring data quality generally is pharma sponsors' key pain point, followed closely by the time required to conduct traditional data review. Despite electronic data capture (EDC) and the availability of electronic medical records (EMR), most clinical trials still rely on

on-site monitoring to verify data and detect quality issues; however, on-site monitoring is slow, costly, and does not provide a holistic view of the data. Additionally, data comes into the trial from multiple sources (e.g., medical records, local/central labs, electronic patient reported outcome [ePRO] systems, central adjudicators, etc.), so reconciling data in the EDC system against these various sources of data always constitutes a challenge. Even when the data is harmonized, uncertainty often plagues clinical trial teams regarding how to analyze or utilize that data in real time.

Part of this uncertainty stems from the length of time needed to clean and monitor data, a process rooted in the antiquated paradigm of trial leadership insisting on source document verification (SDV)/ on-site data monitoring. This leadership demand is rooted in the faulty logic that on-site data monitoring always affects data quality positively.

Consider the following real-life example: a prominent pharma firm was conducting a high-enrolling study that relied heavily on SDV to confirm primary endpoint data before progressing to interim analyses or database lock. However, once analytics were implemented (at the sponsor's request), it was discovered some data was not necessarily accurate, despite being confirmed by a trained eye. The pharma company erred in believing that having someone look at the data ensured its accuracy; or more to the point, the company erred in believing SDV ensured the data was accurately captured.

On this trial, patients were stratified — assigned to a particular treatment group in a blinded study — in real time. Stratification was based on a question in the interactive web response system (IWRS): "Did the patient receive prior biological therapy?" — a seemingly innocuous "yes" or "no" question. But a faulty assumption was made: that the person entering data (and stratifying a patient) knew what constituted a biological agent. This assumption led to multiple mis-stratifications across numerous sites, many of them making the same mistake more than once.

Although remote monitoring activities were implemented to identify these types of errors, once a mis-stratification occurs, it cannot be undone. Additionally, the remote monitoring activities initially implemented relied on manual comparison of data across separate databases, a process prone to error. Once the mis-stratifications became a critical issue (threatening to potentially damage the study's integrity), an advanced algorithm was implemented to programmatically look at the data to identify discrepancies. The algorithm allowed the generation of report outputs that identified actual and potential mis-stratifications, including many overlooked during manual review.

Thus, the stratification step would have been ideal for implementation of analytics, using the patients' prior treatments to decide their treatment group assignment, versus leaving the decision to a human. Or, through use of real-time data capture – more accurately, algorithms that could have been implemented to catch these discrepancies immediately — the issue would have been flagged sooner, leading to earlier intervention.

> **Advanced analytics are critically underutilized in the pharmaceutical industry, particularly in early-stage clinical trials, due to concerns over the investment, in-house expertise, or regulatory scrutiny, among other worries.**

CONQUERING DATA CONTEXTUALIZATION AND ANALYSIS

Clinical trials often encounter difficulty in producing meaningful, accurate data visualizations. A key challenge is the data engineers who program visualization tools do not understand clinical data, and often don't utilize proper data sets when creating visualizations. Vendors hired to provide visualizations rely on clinical trial teams to explain which data must be captured and how, as well as how it should be visualized. A fundamental disconnect exists between the clinical person's understanding of how programming works and their ability to translate the science behind a protocol to a programmer.

A more effective approach would comprise having a subject matter expert (SME) in clinical operations who also understands programming, including its available tools and technologies. However, this powerful combination is not an option for most pharma companies or CROs, either due to a lack of such SMEs or due to the substantial licensing fees associated with many brands providing visualization services. In sum, things are getting lost in translation and few translators are available right now.

Still, clinical trial teams have options, including open-source software platforms that can be downloaded and used by anyone free of charge. Sponsors and their vendor partners already have the necessary data and, with someone who can program, could produce effective visualizations in-house with no need to pay a licensing fee. For example, RStudio is a popular open-source statistical software that can produce visualizations, typically programmed by a statistician or a statistical programmer. But team members with background and training in data science, who know how to use the R programming language (or can figure it out), open endless programming possibilities for clinical trials.

Data analysis, too, suffers from misconceptions: that it can only be performed at set time points or at the end of the study (timing), that such analyses can only be performed by statisticians, and that expensive licensing fees are the only recourse for most companies. Data analyses can be performed using the Python, R, and SQL programming languages across several software platforms. RStudio is attractive for this purpose, in lieu of alternatives that charge a licensing fee (e.g., Tableau), because it is widely used in the scientific community.

RStudio likely is not more widely utilized because clinical trial SMEs do not know how to program in R. They may not know how to clean data using anything more advanced than MS Excel, putting them at a disadvantage when much more powerful and validated tools exist. Python, RStudio, and PostgreSQL are pillars in the scientific and data communities. Despite being free, each is supported and regularly updated with new versions, some including more advanced AI tools — and free does not equal error-prone. Still, wary organizations have several ways to manually confirm outputs from open-source programs before "trusting" the data and analyses they provide.

DE-RISK WITH DEPENDABLE DATA

Risk mitigation is an underlying principle in nearly every clinical activity. In addition to reducing risk, cost, and time investment resultant of miscaptured or inaccurately presented data, advanced analytics and data science can help identify fraud in a clinical trial. Fortunately, such fraud is a relatively rare occurrence, but it can destroy all the data from a particular site.

For example, a recent high-enrolling study (~2,000 patients) included a quality-of-life secondary endpoint, captured electronically, allowing patients to complete their questionnaires at home daily. ISG's data scientist provided a heat map of overall compliance by site, with green indicative of higher compli-

ance and red indicative of less-compliant sites. While overall compliance for the study was between 75% and 80%, the highest-enrolling site stood out with almost 100% patient compliance on the heat map.

It warrants mention that the study had been ongoing for months when the heat map was created, so the near fully compliant site was not an early anomaly. Suspicious, we pulled the raw data set and looked at timestamps associated with the completion of patient diary entries: most entries had been completed within minutes of each other over that period of several months. The statistical likelihood of that happening organically, through patient completion of the diaries, is essentially impossible.

The more likely explanation was that the site was filling out the diaries, a revelation that prompted a site audit. The sponsor was pleased the issue was identified, as such holistic signals are not picked up using traditional monitoring, despite their importance in informing whether a site should be permitted to continue enrolling patients. The audit eventually led to a full QA investigation of the site that revealed numerous other issues. Thus, advanced analytics can pick up signals that point to potential fraud, in addition to trends in quality-of-life improvement or decline and potential data quality issues.

FINAL THOUGHTS

Data gathering, contextualization, and analysis all can benefit from existing — but underutilized — technologies and resources. While the U.S. FDA has indicated a desire[2] to see risk-based monitoring (which may include data analytics and statistical methods) applied to clinical research more extensively, pharma organizations are not rushing to be early adopters. They want to see what competitors

are doing first, but more important, they want to see a company successfully leverage analytics on one or more trials that lead to FDA approval.

It stands to reason that regulators would want to see redundancy in these early submissions: more traditional evidence that objectively confirms the findings of data and analytics cited in the submission. However, those initial sponsors inevitably raise their time, cost, and labor investment by having to create that redundancy. Nonetheless, a trial's clinical operations aspect (i.e., monitoring) often monopolizes 60% or more of its overall budget, not including pass-through costs (e.g., airfare, hotel, and meals for on-site monitors). If that percentage can be reduced even by a small amount, the trial's cost is significantly reduced by minimizing the on-site time typically required by leadership driving the project.

To learn more about the application of advanced analytics and data science to clinical trials — in addition to, or in lieu of, manual data review — visit us at www.inseptiongroup.com.

REFERENCES

1. *De, Sourabh. "Hybrid Approaches to Clinical Trial Monitoring: Practical Alternatives to 100% Source Data Verification." National Center for Biotechnology Information, Medknow Publications Pvt Ltd, July 2011, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3159208/.*
2. *"Guidance for Industry Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring." U.S. Food and Drug Administration, Center for Drug Evaluation and Research, August 2013. https://www.fda.gov/media/116754/download*

ABOUT THE AUTHOR

Nina Anderson, BS, MSLS, is Associate Director of Clinical Science with inSeption Group (ISG). With over 17 years' experience in drug development, her vision is to change the paradigm of clinical research by using advanced analytics and data surveillance techniques to bring safe, effective drugs and devices to market more quickly. Nina possesses therapeutic experience in oncology, myeloproliferative diseases, immunology, psoriasis, and women's health. She earned dual degrees in biotechnology from Thomas Jefferson University, as well as a certificate of data science from Rutgers University. Nina is proficient in Python, RStudio, SQL, and Tableau; she can be reached at nanderson@inseptiongroup.com.

ABOUT INSEPTION GROUP

Unified in our passion to bring therapies and an improved quality of life to patients around the world, we will only provide solutions that increase speed and efficiency in making available life-changing treatments to the patients we serve.

We build the foundation for success by assembling an elite team for every client, every time. Every team member has a mutual passion and an insatiable desire to share in offering hope and making a difference for the patients we serve during their fight for a better, longer life.

We believe that only teams of functional experts, who share personal attributes of excellence and grit, can deliver transparency, precision, and certainty to an industry that is rife with complexity and disappointment.

inSeption GROUP